

Lecture 4: Association Rules

Madhavan Mukund

<https://www.cmi.ac.in/~madhavan>

Data Mining and Machine Learning
August–December 2020

Association rules

- Given sets of items I and transactions T , with confidence χ and support σ , find all valid association rules $X \rightarrow Y$
 - $X, Y \subseteq I, X \cap Y = \emptyset$
 - $\frac{(X \cup Y).count}{X.count} \geq \chi$
 - $\frac{(X \cup Y).count}{M} \geq \sigma$
- For a rule $X \rightarrow Y$ to be valid, $X \cup Y$ should be a frequent itemset
- Apriori algorithm finds all $Z \subseteq I$ such that $Z.count \geq \sigma \cdot M$

Association rules

Naïve strategy

- For every frequent itemset Z
 - Enumerate all pairs $X, Y \subseteq Z, X \cap Y = \emptyset$
 - Check $\frac{(X \cup Y).count}{X.count} \geq \chi$
- Can we do better?
- Sufficient to check all partitions of Z
 - If $X, Y \subseteq Z, X \cup Y$ is also a frequent itemset

Association rules

- Sufficient to check all partitions of Z
- Suppose $Z = X \uplus Y$, $X \rightarrow Y$ is a valid rule and $y \in Y$
- What about $(X \cup \{y\}) \rightarrow Y \setminus \{y\}$?
 - Know $\frac{(X \cup Y).count}{X.count} \geq \chi$
 - Check $\frac{(X \cup Y).count}{(X \cup \{y\}).count} \geq \chi$
 - $X.count \geq (X \cup \{y\}).count$, always
 - Second fraction has smaller denominator, so $(X \cup \{y\}) \rightarrow Y \setminus \{y\}$ is also a valid rule

Observation: Can use apriori principle again!

Apriori for association rules

- If $X \rightarrow Y$ is a valid rule, and $y \in Y$,
 $(X \cup \{y\}) \rightarrow Y \setminus \{y\}$ must also be a valid rule
- If $X \rightarrow Y$ is **not** a valid rule, and $x \in X$,
 $(X \setminus \{x\}) \rightarrow Y \cup \{x\}$ **cannot** be a valid rule
- Start by checking rules with single element on the right
 - $Z \setminus z \rightarrow \{z\}$
- For $X \rightarrow \{x, y\}$ to be a valid rule, both
 $(X \cup \{x\}) \rightarrow \{y\}$ and $(X \cup \{y\}) \rightarrow \{x\}$ must be valid
- Explore partitions of each frequent itemset “level by level”

Association rules for classification

- Classify documents by topic
- Consider the table on the right
- Items are regular words and topics
- Documents are transactions — set of words and one topic
- Look for association rules of a special form
 - $\{\text{student, school}\} \rightarrow \{\text{Education}\}$
 - $\{\text{game, team}\} \rightarrow \{\text{Sports}\}$
- Right hand side always a single topic
- **Class Association Rules**

Words in document	Topic
student, teach, school	Education
student, school	Education
teach, school, city, game	Education
cricket, football	Sports
football, player, spectator	Sports
cricket, coach, game, team	Sports
football, team, city, game	Sports

Summary

- Market-basket analysis searches for correlated items across transactions
- Formalized as association rules
- Apriori principle helps us to efficiently
 - identify frequent itemsets, and
 - split these itemsets into valid rules
- Class association rules — simple supervised learning model